Learn with Surprize from a Robot Professor

Anna-Maria Velentza[1,2], Nikolaos Fachantidis[1,2] ,Ioannis Lefkos[1,2]

[1]School of Educational and Social Policies, University of Macedonia, GR

[2] Laboratory of Informatics and Robotics in Education and Society (LIRES), University of Macedonia, GR

***Abstract*— In the years to come social robots will be extensively used in the teaching field. Consequently, it is important to determine how these robots can optimally interact with students. This paper specifically looks at the performance of social robots in place of university professors, in the field of engineering, measuring three outcomes, enjoyment (questionnaire), the acquisition of knowledge (academic test after class and final exam grades) and level of surprise (facial expressions monitored by cameras). Inspired by previous cognitive studies which have evidenced that feelings of surprise and familiarity affect the learning process, the hypothesis of this research is that robot-tutors will increase student's enjoyment. Leading the research to question whether higher levels of enjoyment strongly correlate with a higher acquisition of knowledge through a series of three experiments. Using the same subject material, three variables were tested: 1. one robot-tutor lecture, 2. one human-tutor lecture, 3. two robot-tutor lectures. Initial findings showed that students lectured by a human-tutor once, acquired a higher level of knowledge compared to those lectured by a robot-tutor once. However, those in the robot-tutor lecture shared higher levels of enjoyment as well as increased expressions of surprise. When students were lectured by a robot-tutor for the second time, results showed both a higher level of acquired knowledge and enjoyment and a drop in expressed surprise, compared to both, one human-tutor lecture and one robot-tutor lecture or one robot-tutor lecture. We conclude that high levels of surprise distracted students initially but once overcome, acted as motivation to study.**

***Index Terms*—human- robot interaction, robot university professor, learning outcome, knowledge acquisition, level of enjoyment**

I. INTRODUCTION

The use of social robots is spreading in many aspects of our lives from healthcare to the educational field and beyond. Education needs to prepare for this future by helping students use cutting edge technologies and learn how to interact with robots. Many researchers and teachers around the world have started using social robots in the classrooms in a variety of applications for different age groups. The four most common uses for robots in classrooms are as a) teachers, b) teacher's assistants, c) student's assistants, d) student's peers, or companions. Recent studies showed that social robots, especially from the position of a tutor or peer learner, can be effective at increasing cognitive and effective outcomes in students and have similar outcomes with those of human tutors on restricted tasks (Belpaeme et al., 2018). Social robots, such as the Kaspar humanoid robot, can also be used with the role of a student or pupil to which typical children and children with autism can improve their own understanding of the task and their social coordination with other children by teaching the robot how to replicate such behaviors (Zaraki et al., 2020).

A widely used social robot in education is the humanoid robot 'NAO' which has been programmed to provide games and activities in order to support fun learning in kindergarten (Alkhalifah et al., 2015) and teach a second language to school students (Meghdari et al., 2013). NAO has delivered positive effects in student's learning outcomes and engagement compared to a human teacher (Preau et al., 2019). Robots have also proved to be a beneficial tool for students to understand a broad range of engineering disciplines and have a significant advantage for University level students to develop Computational Thinking and learn about robotics and practice in control principles and programming (Keane et al., 2016).

In this study, we investigate the use of social robots in the university classroom comparing a) a human vs. a social robot tutor and b) the first-time exposure vs the repeated exposure of university students in a robot-tutor in a real classroom environment. Measuring students' learning outcomes and enjoyment levels. The subject of the course focused on basic engineering principles, and the students participating did not have engineering backgrounds. During the different classroom conditions and the repeated exposure of students in a robot-tutor, we examined the role of students' previous experience and surprise in the learning process provided by a robot.

The key findings of the study are: a) when the participating students had a course with the robot- tutor for the first time, enjoyment levels increased but the level of knowledge gained was less compared to the students with a human-tutor. B) On the other hand, as familiarisation increased in the robot-tutor group experienced two robot-tutor lectures had a higher level of knowledge compared to the students with one lesson with a human-tutor and

one lesson with a robot-tutor. And c) those who had a course with the robot-tutor multiple times had greater scores in the final exam in comparison with all the other students. Finally, based on the findings, the study supports the use of robots for teaching basic engineering principles to students without an engineering background at a university level.

II. Related Work

Typically, research into robots in teaching positions focus on the student's learning outcome and attitudes towards robots (Belpaeme et al., 2018). However, there is a gap in research on how to apply educational theories such as the effect of surprise and the effect of familiarity in the learning process successfully in the robots as educators.

Several studies (Verner et al., 2016), (Polishuk & Verner, 2018) propose the use of the humanoid RoboThespian as a robot teacher for school students aged between 11 and 13 years old. A sample of 189 students participated in a science class that included theoretical and practical lessons with the RoboThespian in the role of the tutor, acting based on teleoperated and pre-programmed behaviours. It is worth mentioning that all possible situations cannot be programmed in advance, and in case a student produces a novel reaction for the robot, there would be a too high working load to rely on having a human to teleoperate the robot's response and thus enabling social robots with online learning abilities is proposed(Khamassi et al., 2011, Khamassi et al., 2018). The knowledge quiz given at the end of the class proved that most of the students understand the main concepts of the lesson (Polishuk & Verner, 2018). Despite that fact, students felt 'psychological distance from the robot'(Verner et al., 2016).

Moreover, children aged 8-12 years old in the Netherlands, have been successfully taught about chronic conditions during health education and theoretically understood conditions such as Type 1 Diabetes with the aid of games and quizzes demonstrated by a robot adaptive to the user needs (Henkemans et al. 2013). Similarly, sixth -grade class students in an elementary school in Tokyo were successfully taught about 'Levers' by the android robot SAYA during a one-hour science lesson(Hashimoto et al., 2013).

You(You et al. 2006) used the Robosapien V1 as an assistant in English class and realised that students paid less attention to the robot during the second week when it asked them to be attentive. Xu(Xu et al., 2014) used the NAO robot as a lecturer in a university class for 30 minutes and evaluated student's learning outcomes with a quiz related to the course. The robot performed two different mood states (positive and negative) but students failed to identify this difference and there wasn't a statistically significant difference between the students' correct answers in the two conditions. On the contrary, Kennedy and his colleagues, when comparing the learning outcome of students learning with a social robot- tutor compared to a human one, found that students with the human- tutor had greater learning acquisition(Kennedy et al., 2016). In another research, university students were asked to evaluate the ideal characteristics of a robot to perform as a university-tutor, and then they witnessed an actual course with the Nao robot and re-evaluated. Results showed that they had significantly different opinions and expectations before and after having a course with the robot-tutor(Velentza, Pliasa, Fachantidis, 2020).

In another study, the less humanoid robot Baxter was stressed against a human teacher by teaching computational concepts in students from 6 to 16 years old. Results showed that students' opinion about their class experience before and after the lecture mainly varied according to student's age(Fernandez-Llamas et al. 2018). Conde et al(Conde et al., 2016) also used the Baxter robot to teach Kindergarten to 18 years old students. Despite the robot's appearance, most of the students felt comfortable interacting with it, while the youngest students reported becoming friends with the robot. Robots can help students to have more positive learning attitudes and less anxiety since they can be less emotive, in comparison with a human- teacher (Alemi et al., 2014). Kwon (2015) asked students aged between 12 and 17 years old in a Hong Kong elementary school if they prefer a robot against a human teacher and why. The answers were equally shared between the two options, with the students who preferred the robot teacher highlighting the opportunity for a more personalised experience based on their individual needs and without punishment. Serholt et al (2017) conducted a series of cross-country focus groups between practicing and pre-service teachers who were enrolled in master studies in Education, discussing their beliefs on the use of educational robots. Results showed that although none had previous experience with teacher robots, all were generally positive in their use.

There are contradictory findings regarding the effectiveness of a robot-tutor in comparison with a human-tutor regarding the students learning outcome(Preau et al., 2019), (Xu et al., 2014), (Kennedy et al., 2016). It is important to examine the reason behind this inconsistency and more importantly to find out the key factor that

affects the students'- robot tutor interaction. Since the use of robots in the role of a tutor is relatively new, our source of theories will be based on the field of Psychology and Learning studies.

Psychological experiments from the early studies in learning in 1960, showed that familiarity is a crucial factor in the learning procedure(Epstein et al., 1960). Students who feel familiar with the group they are working in, have a faster rate in learning tasks in comparison with those who feel unfamiliar. Familiarity leads to the development of psychological safety and thus increase the learning outcome(Staats et al., 2010). A meta-analysis in 2016 indicated that psychological safety influences task performance and learning(Frazier et al. 2017). On the other hand, Holland et al used transient inactivation techniques showing that the surprising omission of an expected event may enhance attention to stimuli that remain present, such that subsequent learning about those stimuli is enhanced (Holland & Gallagher, 2006). Additionally, facial expressions of surprise are strongly positively correlated with informational and motivational effects(Reisenzein et al., 2019)

Based on those findings, we understand that although surprise and psychological safety/ familiarity seem contradictory, both enhance the learning outcome. The surprise effect is important to be controlled in order not to end up distracting the students' attention. In a comparison between gaining knowledge and level of enjoyment in visitors to a museum after having a tour from one or two guide robots, visitors remembered more from the one robot but enjoyed more the tour from the two collaborative robots. It is possible for the visitors to pay attention to the surprising target which is the two robots and thus, they did not explicitly pay attention to what the robot was saying (A.-M. Velentza et al., 2019). Familiarity should also be balanced since there is evidence that students can get bored (Zhen-Jia You et al., 2006). The use of robots in the educational procedure is relatively new and thus we need to re-consider all those factors in the new context of robot-tutors. Besides, for university students, it is not only important to gain knowledge, but also to enjoy the course (James & Nerantzi, 2019).

III. PRESENT STUDY

Based on the evidence defined in the Introduction section, the present study investigates a university teaching procedure with a robot professor, measuring how this can affect student's learning outcomes and enjoyment of the course. Our goal is to teach basic engineering principles to first-year university students without engineering background in an educational department with the help of a humanoid robot professor. Teaching engineering principles has proven successful with the use of robots(Keane et al., 2016).

We conducted three sets of experiments. First of all, we followed the experimental set up used in previous studies (Xu et al., 2014), (Kennedy et al., 2016), (Fernandez-Llamas et al. 2018) and investigated whether there are any differences in a class handling by a human or robot- tutor. Hence, we compared the effect of the nature of the tutor on the learning outcome, as measured by a learning quiz (LQ) and the level of enjoyment measured by a relevant questionnaire (JQ) of the students during the class. Secondly, we compared the learning outcome and enjoyment level in students who had a course with a robot-tutor for the first time and for the second time, by conducting a second robot-tutor teaching course. In addition to the LQ and JQ scores, the courses were filmed and analysed by an expert psychologist to recognize the feeling of surprize from the students' facial expressions during the different courses. Finally, we compared the students' long-term learning outcome during the final exams by comparing the scores of those who had two courses with the robot-tutor, one course, or no course at all.

A. Hypothesis

The hypothesis of this research is that robot-tutors will increase student's enjoyment. As for the learning outcomes, it is expected that students will remember more information from the course and thus pay more attention when they have a human tutor in comparison with the robot tutor (in the first set of experiments).

More specifically, in the first experiment the study investigates whether, in a class of first- year students with no engineering background, the students will acquire more knowledge from a human- tutor in comparison with a robot- tutor, and if the nature of the tutor plays a role in the student's level of enjoyment. If the hypothesis is correct, the robot-tutor will trigger higher levels of surprise to the students, based on the facial expression analysis, leading to less correct answers in the memory test than the human-tutor. Although the university is a new experience and the experiment conducted on the first day of university, students are more familiar with human tutors. On the other hand, we expect that students in the robot condition will get higher scores in the JQ.

In the second experiment, we utilise, a second round of the first experiment, where the robot- tutor performed the lesson for all the students. In this case, it is expected that students who experience the robot for the second time, will have higher learning acquisition because of increased familiarity with a robot-tutor, while those who experience the robot for the first time, will feel surprised and will have less correct answers in the memory test (i.e. they will recall less information from the lesson). Moreover, in this second experiment we expect that after all students experience a class with a robot teacher, they will have similar scores in the JQ or based on You's findings, those who have the course for the first time to have higher JQ (Zhen-Jia You et al., 2006). Finally, at the final exam, it is expected that students who experienced the robot- tutor two times, will be more motivated (Reisenzein et al., 2019) and achieve higher scores in comparison with the other students.

IV. EXPERIMENTAL STUDIES

A. Experiment 1: Human vs Robot Tutor

a) Participants

In the first experiment, 138 people participated, 7 Males and 131 Females, aged 18-28 years old. They were all freshmen, undergraduate students, in the School of Educational and Social Policies, on their first day in the mandatory course "Basic Principles of Information and Communication Technologies". To avoid spoiling and/or developing hype from the robot lectures, we decided to experiment with the human tutor before the one with the robot. The experiments took place on two consecutive days, in the same hour and in the same classroom. The first group participated in the first experiment with the human-tutor (N=66) and the second group in the class conducted by the robot-tutor (N=72); all with normal or corrected to normal vision and hearing and the language of the course was their native.

b) Design and Procedure

The experimental design was between-participants (i.e. having different participants in each). During the first day, the course was conducted by the course's professor (human- tutor), who has taught the same as well as similar subjects for 20 years. During the second day, the course conducted by the Aldebaran Nao V3.3 (robot-tutor) which was programmed to do similar body movements with the human-tutor, since humans tend to trust more and feel more familiar when they are close to an agent/ system that has similar characteristics with them(Verberne et al., 2015). The tutors, human or robot started introducing his/itself (based on the instructions of Rainer and Galan 2010) and started to explain the guidelines of the course. The script was the same in both conditions. The tutors, human, and robot stood in the middle of the class, as shown in Figure 1, and taught about basic principles of Cryptography. Both tutors used the same presentation and teaching style while the duration of both classes was 30 minutes. Moreover, since the students did not know their classmates yet, there were two undercover researchers in the audience to answer the tutor's questions. Especially in the robot tutor condition that seems to make people trust the robot more (Lomas et al., 2012) since it described the researchers who raised their hand with detail (i.e. the girl in the second row of seats with the blue t-shirt) and gave accurate feedback about their reply.

After the end of the class, the tutor thanked the students for their attention and a teaching assistant gave them firstly the LQ, and then the JQ and a demographic and familiarity questionnaire. The participants had as much time as they wanted to fill them in anonymously. Three cameras were recording the lecture, as can be seen in Fig.1, (one on the left side, one on the right, and one standing on the desk). After the end of the lecture in all condition's participants signed that they agree with the use of their data for research purposes and they had a one-month period to notify the teaching assistant to dispose of them.

FIG.1: CLASSROOM SET UP IN THE ROBOT CONDITION

c) Measurements

In the learning quiz (LQ) there were a total of 10 questions, 5 Multiple Choice Questions (MCQ) with four multiple answers and only one correct answer per question and 5 open questions, since open-ended and multiple-choice questions measure different aspects of comprehension processes(Ozuru et al., 2013). The questions were about the content of the lecture. The LQ was marked by the teaching assistant who was teaching the course together with the professor for already three years, with an exam marking scheme, in order to avoid biases from the tutor who may manage to relate the questionnaires with the experimental condition.

To evaluate the student's level of enjoyment (JQ) from the course, we used the 35 items' Likert scale questionnaire that has been used by Velentza et her colleagues (2019) to evaluate students' impressions after the robot's storytelling. For the use of the JQ questionnaire we asked and received permission from the copyright holder. For the demographic characteristics (i.e. gender, age) and familiarity questions, we created a questionnaire with seven Likert scale questions regarding the participants' relationship with technology, robots, and the subject of the course e.g. "have you ever used a robot". All the questionnaires were anonymous and after they got collected, they were marked with a unique code per person.

d) *Data Analysis*

We analysed the LQ scores for the correct answers with a paired sample t-test (Montgomery et al., 2009). Additionally, based on Mutlu, Forlizzi, and Hodgins (2006) research findings, explaining that the highest learning outcome does not mean that the students enjoyed the learning process equally, we investigated whether the experimental condition that leads to higher LQ scores, equitably leads to high scores in JQ with the use of an ANOVA test. To analyse the students' scores of the JQ, we handled a between-participants Bonferroni multiple comparison test to find if there were any significant differences between the different conditions from the JQ responses. We used a multiple comparison analysis because we also compared the students' scores from the first and the second experiment. We also applied a Pearson r correlation analysis, on the demographics, LQ and JQ for both robot and human conditions. All the analyses were performed with the aid of SPSS 25.

For the demographic and familiarity data, we performed a factor analysis to split the questions into groups. By running the Factor Analysis, we end up with 3 categories. In the parenthesis we report the Cronbach's alpha reliability measure for each category: 1) how they appreciate the use of technology in education (α=.85), 2) if they are familiar with the course's subject (α=.89), 3) if they have previous experience with robots (α=.91). Data analysis remains the same for all Experiments.

For the video analysis, we used the ELAN 5.9 Software(ELAN, 2020) to measure the surprise effect through the students' facial expressions. The analysis was conducted by a cognitive psychologist, non-biased from the experimental design. The videos were muted without revealing the tutors' spot to avoid any recognition of the tutors' nature (robot or human). The facial expression of surprise came upon the theory of evolution and the evolutionary emotion psychologists(Velentza, 2020). The physiological reaction is marked by skin conductance increase, raised eyelids and eyebrows which are not drawn together, dropped down jaw, and pupil dilation. Additionally, head movements and in many cases bringing the hands up to shield the face (Reisenzein et al., 2019). For the analysis, we took into consideration only the students who were totally visible during the whole procedure. The analysis started when the tutor welcomed the students and finished the moment the teaching assistant gave the first questionnaire to the first student. We created one tier per participant in which we marked the milliseconds during which a surprise motion appeared. At the end of the procedure, we summed up the marked milliseconds for all the tiered participants and we divided them with the number of participants per condition to end up with one number, the average milliseconds of surprise per condition. Additionally, we applied a t-test to determine any statistically significant differences in the students' level of surprise between the conditions.

e) *Results*

The JQ score of students in the human-tutor condition (Mean Values (*MV*)=3.41, Standard Deviation (*SD*)= 0.65) was lower than the score of students in the robot-tutor condition (*MV*=3.78, *SD*= 0.654). The Bonferroni analysis confirmed that this difference was statistically significant at $p<.001$ (d=13.75, Std. Error=3.04, Lower Bound=5.67, Upper Bound=21.82), as depicted in Fig.2

The maximum score in the LQ was 25, 5 for the multiple choice and 20 for the open questions. In the human-tutor condition, the student's scored higher (*MV*=9.29, *SD*=2.8), 37,16% correct answers than in the robot-tutor condition (*MV*= 8,2, *SD*=2.5), 32,8% correct answers. The difference proved to be significant at (*t(70)= 2.37, p= .019, d=1.08*)[1]. Similarly, the scores in the multiple-choice questions in the human-tutor condition (*MV*= 3,42, *SD*=1.1) and in the open questions, (*MV*= 5,9, *SD*=2.1) were higher than multiple-choice questions, (MV= 3,04, SD=1.05) and open questions (*MV*= 5,16, *SD*= 2.1) in the robot-tutor condition, which can be found detailed in Fig.3.

The ANOVA analysis indicated that the students with greater learning acquisition (higher scores in the LQ) did not have equally high scores in the JQ.

The correlation analysis between the students' JQ and LQ scores showed that in both conditions, students who were familiar with robots had higher scores in the LQ and thus higher learning acquisition, r(65)=.514, p=.<001, in human-tutor condition and r(71)=.549, p=.<001 in robot-tutor condition, as shown in Figure 2. There were no correlations between the rest of the categories and the questionnaires' scores, as shown in the supplementary Figures, Annex Section.

Finally, students in the robot- tutor condition had higher average time of surprise based on their facial expressions (*MV*=168.087ms, *SD*=3.94) in comparison with the human-tutor condition (*MV*=18.09ms,*SD*=2.025), as indicated by t-test (t(25)= -5.07, p= <.001, d=2.45).

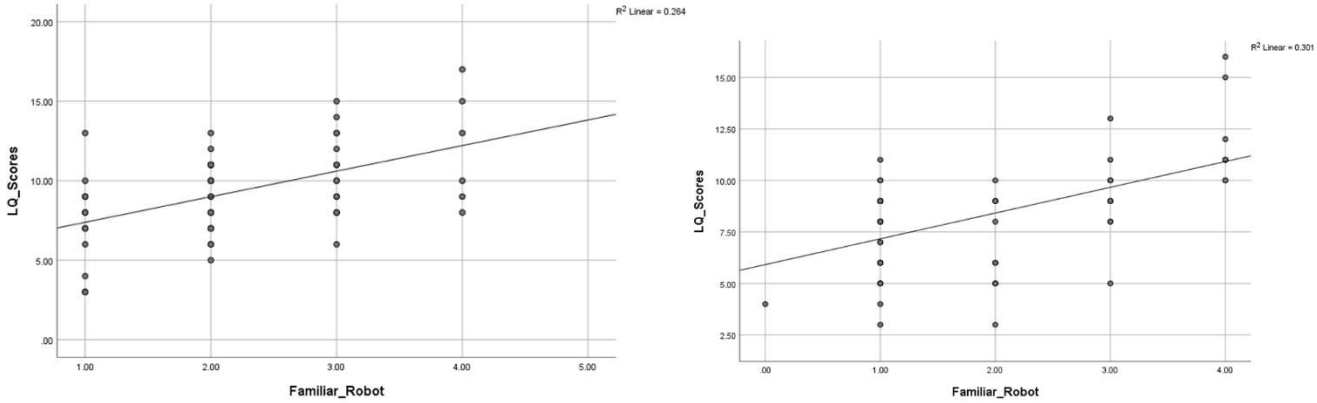

FIG.2: CORRELATION BETWEEN STUDENTS' FAMILIARITY WITH ROBOTS AND LQ SCORES. LEFT: HUMAN-TUTOR CONDITION, N=66. RIGHT: ROBOT-TUTOR CONDITION, N=72

f) Discussion I

The students had higher knowledge acquisition from the human- tutor lesson, as found form (Kennedy et al., 2016) and thus paid more attention to the human tutor. Based on the hypothesis, students are familiar with human-tutors and thus they experienced the familiarity effect in the human-tutor condition that was missing from the robot-tutor. On the other hand, students in the robot-tutor condition experienced the surprising effect (that was missing from the human-tutor condition as shown also from the average surprise time per condition) which based on (Holland & Gallagher, 2006) must be balanced between 'surprised' to attract their attention but not 'too surprised' to distract them. Despite the feeling of surprise that lead to lower learning outcomes in comparison with the human tutor, we believe that this effect in the robot-tutor condition resulted in higher JQ scores, meaning that students enjoyed the class more.

B. Experiment II: 1st time vs 2nd time robot tutor

a) Participants

The total number of participants was 89, 6 Males and 83 Females, aged 19-28 years old. They all had participated in the 1st experiment, either in the robot or the human condition. In the first group who had a course with a robot-tutor for the second time, N=37, and in the second group who had a robot-tutor for the first time, N=52. We attribute the smallest sample of students in comparison with the 1st experiment, to the fact that the lecture took place one week before the Christmas break.

[1] t(freedom degrees)=tscore, p = pvalue (statistically significant at ≤.05), d=mean differencestandard deviation

b) Design and Procedure

The experimental design was between-participants and both the groups had a lecture with the robot- tutor. For the first group (Rob2) it was their second robot-tutor course while for the second group (Rob1) it was their first. The lecture was conducted 10 weeks after Experiment I and it was about the internal and external systems of a computer, the storage devices, and social issues about technology such as technological illiteracy and lasted for 30 minutes. The design and procedure were similar to the first experiment.

c) Data Analysis

The data analysis was similar to Experiment I, except for a small adjustment in the video analysis. Some of the students had changed their timetables and thus it will be impossible for the psychologist to exclude the students who did not belong in the group. For this purpose, at the beginning of the course, the robot asked the students who were enrolled on a different day to raise their hands. Those students were excluded from the analysis (i.e. in the group that was supposed to have a course with the robot-tutor for a second time, the robot ask the students to raise their hands those who had never had a course with a robot).

d) Results

The mean JQ score in those who had a lecture with the robot for the first time (MV= 3.83, SD=0.58) was similar to those who had the lecture for a second time (MV= 3.92, SD=0.5). The Bonferroni analysis indicated the similarity at p=.882 (d=2.87, Std. Error=3.89, Lower Bound=-7.19, Upper Bound=12.94).

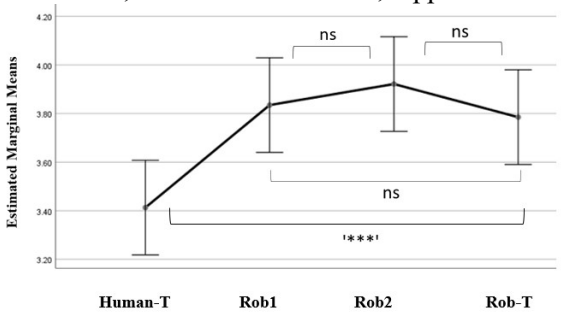

FIG.3: MEAN VALUE OF JQ SCORES, HUMAN-T AND ROB-T FROM EXPERIMENT I AND ROB1 AND ROB2 FROM EXPERIMENT II

The LQ scores for Rob1 (*MV*=11.73, *SD*=4.88), *47% correct answers*, were statistically significant lower in comparison with Rob2 (*MV*=15.04, *SD*=4.49), *60.16% correct answers* as proven from the t-test analysis, t(44) = -3.25,p= .0016,d=3.31. Similarly, the scores in the multiple-choice questions in the R1 (*MV*= 2.98, *SD*=1.4), *59.6% correct answers* and in the open questions, (*MV*= 8.75, *SD*=4.5), *43.75% correct answers* were lower than the multiple-choice questions, (*MV*= 3,35, *SD*=0.95), 67% correct answers and the open questions (*MV*= 11.68, *SD*= 4.41), *58.4% correct answers* in the R2, which can be found detailed in Fig.4.

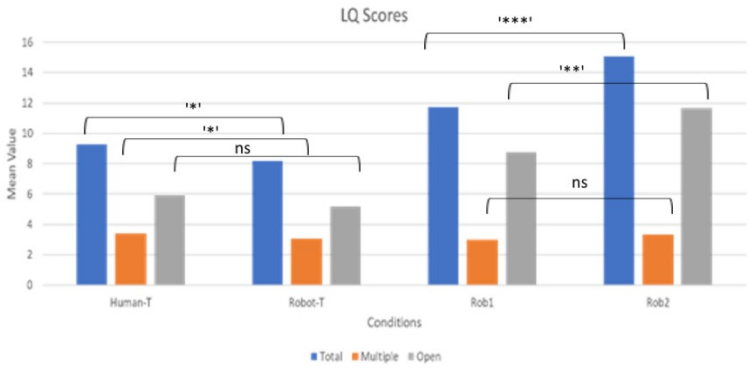

FIG.4: MEAN VALUE OF LQ SCORES

The students in both Rob1 (p=.59) and Rob2 (p=.094) conditions did not have equally high LQ and JQ scores. Additionally, there were no correlations between the LQ and JQ scores and the familiarity questions.

The facial expression video analysis indicated that the average time of surprise in the students who participated in the Rob1 (*MV*=53.22ms, *SD*=3.46) was lower than those who participated in the Rob2 condition (*MV*=58.17ms, *SD*=3.12), as indicated by t-test (t(25)=2.67, p=.008,d=1.06).

e) Discussion II

Consistent to the hypothesis, those who had a second course with the robot tutor had statistically significantly higher scores in the LQ and thus higher learning acquisition than those who had a lesson for the first time. As for the level of enjoyment, as shown from the JQ scores, students in both conditions had similar scores, since they both had a course with a robot-tutor. Regarding the video analysis, we expected a higher average time in those who had a course with the robot-tutor for the first time but instead, the group who had a course with the robot for a second time had higher scores. We believe that the students who first had a course with the human-tutor and then with the robot-tutor were biased from their classmates' opinions who had two times a course with the robot-tutor. It has already been mentioned from Velentza et al., (2020) that college students' expectations about a robot's appearance, functionality, use etc significantly vary from their opinion after witnessing a lectured performed by the Nao robot. Their peers did not have any expectations since the first time they did not know that they are going to have a lesson with a robot-tutor and the second time, based on their previous experience they knew what to expect. On the other hand, those who had a lecture with the human-tutor, heard the stories from their classmates and formed high expectations -since those who had a lecture with the robot-tutor had high average surprize time- and they probably felt disappointment.

V. Comparison of Experiment I vs Experiment II

The level of students' enjoyment was higher in the Rob2 condition, where the students had a course with the robot-tutor for the second time, as indicated by the Bonferroni analysis, in comparison with Rob1 and both the human-tutor (p=.<001, d=18.46, Std. Error=3.67, Lower Bound=8.68, Upper Bound=28.25) and robot-tutor condition (p=.571, d=4.71, Std. Error=3.66, Lower Bound=-4.7, Upper Bound=14.18) in Experiment I. Between Experiment I (where half of the students had a lesson with a human tutor while the others with a robot) and Experiment II where all students had a lesson with a robot tutor, we found out that students who had a lesson with a robot tutor evaluated higher their experience from the lesson.

As for the LQ scores and thus to the student's level of knowledge acquisition and attention during the lesson, even though we cannot directly compare the scores from the two experiments, since the lessons were on different subject matter and also had different previous experience regarding the tutors, we noticed that students paid more attention to the lecture when they were familiar with the tutor. In Experiment I, students who had one lesson with the robot tutor and one with the human tutor had 4.36% more correct answers comparing with the students who had a robot tutor while in Experiment II, students who had a robot tutor for a second time had 13.16% more correct answers in comparison with the students who saw a robot tutor for the first time.

We also compared the students' level of surprize within groups. Those who had a course with the human-tutor and then with the robot-tutor, had as expected statistically significantly higher average time when they witnessed the robot-tutor lesson, as indicated by t(25)=-4.62,p=<.001,d=2.07. On the other hand, the level of surprise in those who had a course with the robot-tutor for the first time statistically significantly fallen after having a course with the robot-tutor for a second time (t(25)=3.84, p=<.001,d=1.44). Those results indicated that, as expects, students were more surprized when they had a course with a robot- tutor for the first time.

Students' level of surprise in Experiment I in the robot-tutor condition was 8.3 times higher than in the human-robot condition. As expected, the students in Rob2 were more surprised than the students participating in the human-tutor condition. On the other hand, the Rob1 scores of the average time were not similar to the scores of the students in Experiment I as expected. We believe that apart from the feeling of familiarity with the tutor, the familiarity with the course subject plays a role in the result. The students in Experiment II had ten weeks of lessons before having the lesson with the robot. Moreover, we believe that students in Rob1 were not equally surprised with the students in Experiment I because they were informed by their classmates about the robot-tutor which removed some of the element of surprise.

In conclusion, students paid more attention to a robot-tutor after having a second lesson with it. Moreover, they found the lesson more attractive, interesting, and, generally, they evaluated their experience as being significantly better, when the tutor was a robot, in comparison with a conventional human-professor.

C. EXPERIMENT III: EXAM DAY

a) Participants

We separated the students into three categories, (a) those who had one lesson with the robot tutor, N= 78 (b) two lessons N= 60, (c) never, N=64. Those who belong in the third group are students who were missing from the class the days of the experiments.

b) Procedure

The final exam for passing the course was developed through a university software platform and was held in the University's pc room. The students were registered to give the exam in different groups according to the first letter of their surname, every one hour on the same day. Participants from the three previously mentioned categories were mixed into those groups. In the total number of exam questions, we added the 6 more difficult questions -those with the less correct answers- from the LQ from Experiment I and II.

c) Data Analysis

To analyse the data, we applied the Fisher's Least Significance Difference Test (LSD) which is a powerful post hoc comparison for three groups.

d) Results

The maximum score in the six chosen questions of the final test was six. Students who had two courses with the robot-tutor (*MV*=4.36, *SD*=1.14) had statistically significant higher scores than those who had one course with the robot- tutor (*MV*=3.76, *SD*=1.33), as indicated by the LSD at p=.005 (Mean Diff=.6002, Std Error=.21, Lower Bound=.183, Upper Bound=1.01). Those who had never seen a robot-tutor (*MV*=3.3, *SD*=1.37) had statistically lower scores from both the group who saw the robot for one time (p=.031, Mean Diff=.45, Std Error=.208, Lower Bound=.042, Upper Bound=.86) and especially from the group who saw the robot for two times (p=<.001, Mean Diff=.1.05, Std Error=.23, Lower Bound=.59, Upper Bound=1.5), as shown in Fig.5

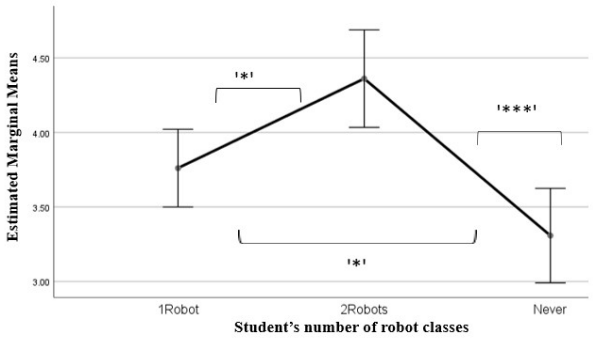

FIG.5: ESTIMATED MARGINAL MEANS OF EXAM DAY TOTAL SCORES

VI. GENERAL DISCUSSION AND CONCLUSIONS

The present study focused on the interaction between first- year university students without engineering backgrounds and a robot-tutor. We were interested mainly in how the robot-tutor can improve the students' knowledge about basic engineering subjects such as the internal and external computer systems as well as the level of enjoyment from the course's experience. The students had one or multiple courses held by a robot-tutor and one course with the robot-tutor and one with the human-tutor, and we also evaluated the level of surprise they experienced during the courses and how this may affect their knowledge acquisition and level of enjoyment. The study was initially motivated by the contradictory outcomes of similar studies about the effectiveness of a robot-tutor.

The first outcome is that the first time the students had a course with a robot-tutor; they retained less knowledge from it, in comparison with the course held by a) a human-tutor and b) a robot-tutor for a second time. In particular, they had lower scores in the given learning questionnaire (LQ) most probably because they were not familiar with a robot-tutor, they spent almost eight times more time feeling surprised and did not pay attention in the lesson but in the robot- tutor itself. Moreover, based on our hypotheses, students had higher knowledge gain

when they felt familiar with the course procedure and the nature of the tutor and thus in Experiment I, when compared a human against a robot tutor, students had higher LQ scores in the human- tutor condition. Respectively for Experiment II, students who had a second course with the robot-tutor had higher LQ scores in comparison with those who had a course for the first time.

Despite this, students in all robot conditions had statistically significantly higher scores in the enjoyment questionnaire (JQ) in comparison with the human-tutor condition. In more detail, they enjoyed the course more even in the cases where their knowledge gain felt low. The JQ score was statistically significantly lower in the human-tutor condition in comparison with all the robot conditions, while the highest JQ score and thus the highest enjoyment level, appeared in the second time students had a course with the robot-tutor (Rob2).

Regarding the camera analysis and the students' surprise as tiered from their facial expressions, as expected, students in all robot conditions had significantly higher average time per condition in comparison with the human-tutor condition. On the other hand, in Experiment II, students did not have an equally high average time when they had a course with the robot-tutor (Rob1) compared with the students in the same condition from Experiment I. Instead, they had a similar average surprise time with the group of students who had a course for the second time (Rob2), but they still had statistically significant lower scores than students in Rob2. They were unfamiliar with the robot- tutor but they were not very surprised because they were informed by their classmates about the robot-tutor which removed some of the element of surprise. Although, this needs further investigation. It is also very promising that students who had a second course with the robot-tutor remained surprised, yet less surprised than in Experiment I where they had their first experience and they also continued to feel more joy from the course.

The scores during the final exam confirmed that the robot's presence motivated the students as shown from the statistically significant higher scores than they had in comparison with those who did not have any course with the robot-tutor. Additionally, those who had a course for a second time had a statistically significant higher score even than those who had one course with the robot-tutor. The students' results in a final exam is a combination of attending classes and systematic reading, and thus we cannot jump to conclusions based on them. Moreover, it is possible that those who never had a course with the other, never attended the classes or miss many lessons and thus they had poor performance in the final exams. Although it seems that the presence of the robot played a key role in their motivation as also stated by Reisenzein et al(2019).

Generally, the results demonstrate that social robots can serve as tutors teaching engineering subjects in students without engineering backgrounds such as teachers and educators. One of the limitations is the gender of the participants who are in mostly females and thus it needs further investigation before generalize the results for all genders. Additionally, it is worth mentioning that we do not know if the students had been taught by robots for many years, they would have felt surprised with the human-tutor, and no surprise with the robot-tutor. So, the results may have been reversed compared to here, and we should keep this into consideration for future research.

As for future work, we encourage researchers to examine the effect of familiarity and surprise in more controlled conditions and experiment with the use of robot- tutor multiple times.

REFERENCES

Alemi, M., Meghdari, A., & Ghazisaedy, M. (2014). Employing Humanoid Robots for Teaching English Language in Iranian Junior High-Schools. *International Journal of Humanoid Robotics*, *11*(03), 1450022. https://doi.org/10.1142/S0219843614500224

Alkhalifah, A., Alsalman, B., Alnuhait, D., Meldah, O., Aloud, S., Al-Khalifa, H. S., & Al-Otaibi, H. M. (2015). Using NAO Humanoid Robot in Kindergarten: A Proposed System. *2015 IEEE 15th International Conference on Advanced Learning Technologies*, 166–167. https://doi.org/10.1109/ICALT.2015.17

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. *Science Robotics*, *3*(21). https://doi.org/10.1126/scirobotics.aat5954

Blanson Henkemans, O. A., Bierman, B. P. B., Janssen, J., Neerincx, M. A., Looije, R., van der Bosch, H., & van der Giessen, J. A. M. (2013). Using a robot to personalise health education for children with diabetes type 1: A pilot study. *Patient Education and Counseling*, *92*(2), 174–181. https://doi.org/10.1016/j.pec.2013.04.012

Conde, M. Á., Fernández, C., Rodríguez-Lera, F. J., Rodríguez-Sedano, F. J., Matellán, V., & García-Peñalvo, F. J. (2016). Analysing the attitude of students towards robots when lectured on programming by robotic or human teachers. *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality*, 59–65. https://doi.org/10.1145/3012430.3012497

ELAN (Version 5.9) [Computer software]. (2020). Nijmegen: Max Planck Institute for Psycholinguistics. Retrieved from https://archive.mpi.nl/tla/elan". (n.d.).

Epstein, W., Rock, I., & Zuckerman, C. B. (1960). Meaning and familiarity in associative learning. *Psychological Monographs: General and Applied*, *74*(4, Whole No. 491), 22–22.

Fernndez-Llamas, C., Conde, M. A., Rodrguez-Lera, F. J., Rodrguez-Sedano, F. J., & Garca, F. (2018). May I teach you? Students' behavior when lectured by robotic vs. human teachers. *Computers in Human Behavior*, *80*(C), 460–469. https://doi.org/10.1016/j.chb.2017.09.028

Hashimoto, T., Verner, I. M., & Kobayashi, H. (2013). Human-Like Robot as Teacher's Representative in a Science Lesson: An Elementary School Experiment. In J.-H. Kim, E. T. Matson, H. Myung, & P. Xu (Eds.), *Robot Intelligence Technology and Applications 2012: An Edition of the Presented Papers from the 1st International Conference on Robot Intelligence Technology and Applications* (pp. 775–786). Springer. https://doi.org/10.1007/978-3-642-37374-9_74

Holland, P. C., & Gallagher, M. (2006). Different Roles for Amygdala Central Nucleus and Substantia Innominata in the Surprise-Induced Enhancement of Learning. *Journal of Neuroscience*, *26*(14), 3791–3797. https://doi.org/10.1523/JNEUROSCI.0390-06.2006

James, A., & Nerantzi, C. (2019). *The Power of Play in Higher Education: Creativity in Tertiary Learning*. Springer.

Keane, T., Chalmers, C., Williams, M., & Boden, M. (2016). The impact of humanoid robots on students' computational thinking. In P. Albion & S. Prestridge (Eds.), *Australian Council for Computers in Education 2016 Conference: Refereed Proceedings:* (pp. 93–102). The Queensland Society for Information Technology in Education (QSITE). http://conference.acce.edu.au/index.php/acce/acce2016/paper/view/39/31

Kennedy, J., Baxter, P., Senft, E., & Belpaeme, T. (2016). Heart vs hard drive: Children learn more from a human tutor than a social robot. *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 451–452. https://doi.org/10.1109/HRI.2016.7451801

Khamassi, M., Lallée, S., Enel, P., Procyk, E., & Dominey, P. F. (2011). Robot Cognitive Control with a Neurophysiologically Inspired Reinforcement Learning Model. *Frontiers in Neurorobotics*, *5*. https://doi.org/10.3389/fnbot.2011.00001

Khamassi, M., Velentzas, G., Tsitsimis, T., & Tzafestas, C. (2018). Robot Fast Adaptation to Changes in Human Engagement During Simulated Dynamic Social Interaction With Active Exploration in Parameterized Reinforcement Learning. *IEEE Transactions on Cognitive and Developmental Systems*, *10*(4), 881–893. https://doi.org/10.1109/TCDS.2018.2843122

Kwok, V. H. Y. (2015). Robot vs. Human Teacher: Instruction in the Digital Age for ESL Learners. *English Language Teaching*, *8*(7), 157–163.

Lomas, M., Chevalier, R., Cross, E. V., Garrett, R. C., Hoare, J., & Kopack, M. (2012). Explaining robot actions. *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, 187–188. https://doi.org/10.1145/2157689.2157748

Meghdari, A., Alemi, M., Ghaazisaidi, M., Tahen, A. R., Karimian, A., & Vakih, M. Z. (2013). Applying robots as teaching assistant in EFL classes in Iranian middle schools. *In Proceedings of the 2013 International Conference on Education and Modern Educational Technologies (Beijing)*, 67–73.

Montgomery, D. C., Runger, G. C., & Hubele, N. F. (2009). *Engineering Statistics, Student Study Edition*. John Wiley & Sons.

Mutlu, B., Forlizzi, J., & Hodgins, J. (2006). A Storytelling Robot: Modeling and Evaluation of Human-like Gaze Behavior. *2006 6th IEEE-RAS International Conference on Humanoid Robots*, 518–523. https://doi.org/10.1109/ICHR.2006.321322

Ozuru, Y., Briner, S., Kurby, C. A., & McNamara, D. S. (2013). Comparing comprehension measured by multiple-choice and open-ended questions. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *67*(3), 215–227. https://doi.org/10.1037/a0032918

Polishuk, A., & Verner, I. (2018). An Elementary Science Class with a Robot Teacher. In W. Lepuschitz, M. Merdan, G. Koppensteiner, R. Balogh, & D. Obdržálek (Eds.), *Robotics in Education* (pp. 263–273). Springer International Publishing. https://doi.org/10.1007/978-3-319-62875-2_24

Preau, V., Tan, T., Lohues, P., & Cappe, P. (2019). The adventures of a humanoid robot in a French classroom. *Australian Educational Leader*, *41*(2), 54.

Psychological Safety: A Meta-Analytic Review and Extension—Frazier—2017—Personnel Psychology—Wiley Online Library. (n.d.). Retrieved 8 September 2020, from https://onlinelibrary.wiley.com/doi/full/10.1111/peps.12183?casa_token=BH3ggaral_IAAAAA%3AMTJoasV8hUYqRooKFldKSH3OvhdTACYmYoqq4AloVeXy2EWASgeduhqR0_xOjPpCWm4KWLt98ZNtLOw

Rainer, J. J., & Galan, R. (2010). Learning to Be a Good Tour-Guide Robot. *2010 Ninth International Conference on Machine Learning and Applications*, 595–600. https://doi.org/10.1109/ICMLA.2010.92

Reisenzein, R., Horstmann, G., & Schützwohl, A. (2019). The Cognitive-Evolutionary Model of Surprise: A Review of the Evidence. *Topics in Cognitive Science*, *11*(1), 50–74. https://doi.org/10.1111/tops.12292

Serholt, S., Barendregt, W., Vasalou, A., Alves-Oliveira, P., Jones, A., Petisca, S., & Paiva, A. (2017). The case of classroom robots: Teachers' deliberations on the ethical tensions. *AI & SOCIETY*, *32*(4), 613–631. https://doi.org/10.1007/s00146-016-0667-2

Staats, B., Gino, F., & Pisano, G. (2010). *Varied Experience, Team Familiarity, and Learning: The Mediating Role of Psychological Safety*. Harvard Business School.

Velentza, A. M. (2020). Putting the Humans in the Middle of the CPS Design Process. In *Heterogeneous Cyber Physical Systems of Systems* (pp. 175–209). River Publishers.

Velentza, A.-M., Heinke, D., & Wyatt, J. (2019). Human Interaction and Improving Knowledge through Collaborative Tour Guide Robots. *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1–7. https://doi.org/10.1109/RO-MAN46459.2019.8956372

Velentza, A.-M., Pliasa, S., & Fachantidis, N. (2020). *Future Teachers choose ideal characteristics for robot peer-tutor in real class environment*. International Conference on Technology and Innovation in Learning, Teaching and Education, TECHEDU2020.

Verberne, F. M. F., Ham, J., & Midden, C. J. H. (2015). Trusting a Virtual Driver That Looks, Acts, and Thinks Like You. *Human Factors*, *57*(5), 895–909. https://doi.org/10.1177/0018720815580749

Verner, I. M., Polishuk, A., & Krayner, N. (2016). Science Class with RoboThespian: Using a Robot Teacher to Make Science Fun and Engage Students. *IEEE Robotics Automation Magazine*, *23*(2), 74–80. https://doi.org/10.1109/MRA.2016.2515018

Xu, J., Broekens, J., Hindriks, K., & Neerincx, M. A. (2014). Effects of bodily mood expression of a robotic teacher on students. *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2614–2620. https://doi.org/10.1109/IROS.2014.6942919

Zaraki, A., Khamassi, M., Wood, L. J., Lakatos, G., Tzafestas, C., Amirabdollahian, F., Robins, B., & Dautenhahn, K. (2020). A Novel Reinforcement-Based Paradigm for Children to Teach the Humanoid Kaspar Robot. *International Journal of Social Robotics*, *12*(3), 709–720. https://doi.org/10.1007/s12369-019-00607-x

Zhen-Jia You, Chi-Yuh Shen, Chih-Wei Chang, Baw-Jhiune Liu, & Gwo-Dong Chen. (2006). A Robot as a Teaching Assistant in an English Class. *Sixth IEEE International Conference on Advanced Learning Technologies (ICALT'06)*, 87–91. https://doi.org/10.1109/ICALT.2006.1652373

ANNEX

CORRELATIOS LQ-FAMILIARITY QUESTIONS HUMAN-TUTOR CONDITION (N=66)

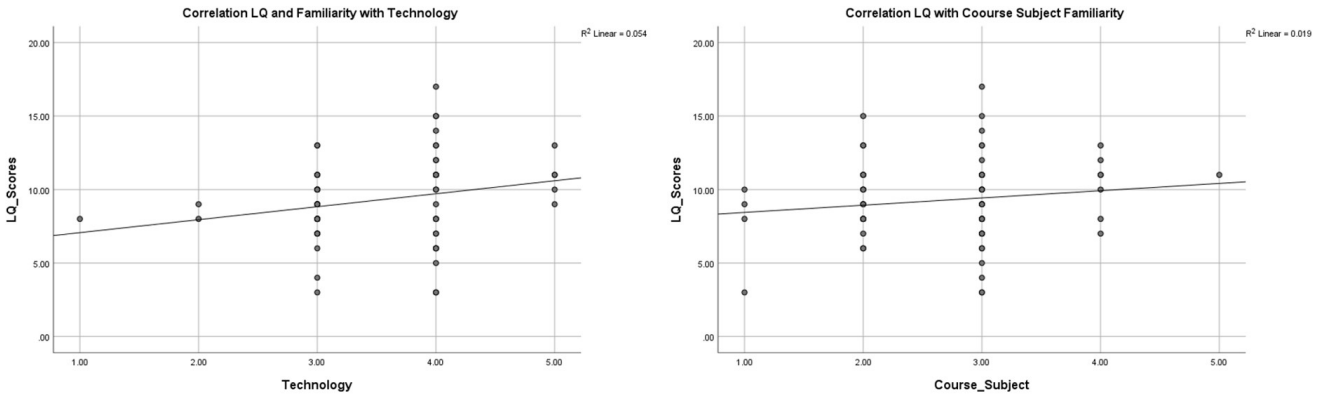

FIG.6: HUMAN-TUTOR CONDITION CORRELATIONS BETWEEN LQ SCORES AND FAMILIARITY WITH TECHNOLOGY AND COURSE SUBJECT

CORRELATIOS LQ-FAMILIARITY QUESTIONS ROBOT-TUTOR CONDITION (N=72)

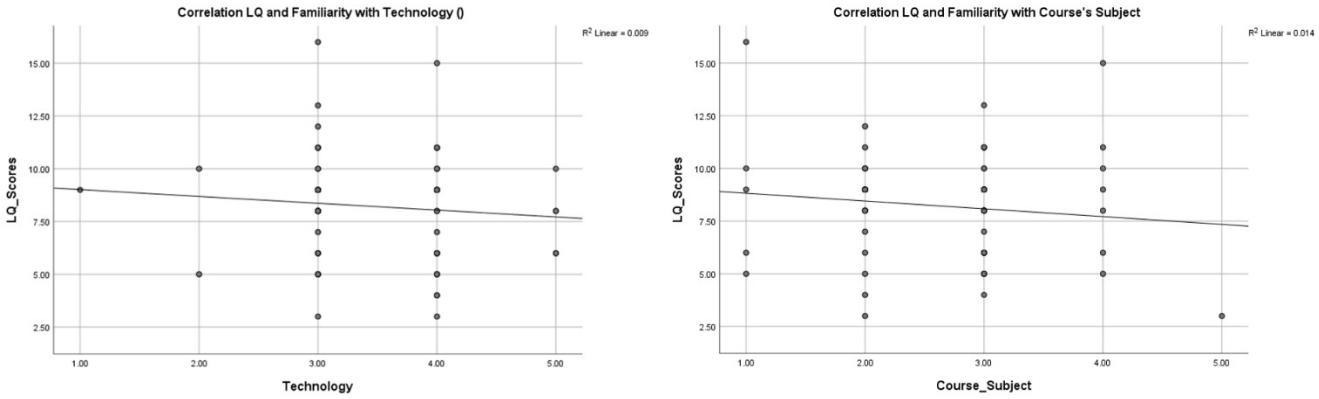

FIG.7: HUMAN-ROBOT CONDITION CORRELATIONS BETWEEN LQ SCORES AND FAMILIARITY WITH TECHNOLOGY AND COURSE SUBJECT